

# Fitting piecewise linear regression functions to biological responses

ELISABETH VIETH

Max-Planck-Institut, W.G. Kerckhoff-Institut (Arbeitsgruppe Statistik),  
6350 Bad Nauheim, Federal Republic of Germany

VIETH, ELISABETH. *Fitting piecewise linear regression functions to biological responses*. J. Appl. Physiol. 67(1): 390–396, 1989.—An iterative approach was achieved for fitting piecewise linear functions to nonrectilinear responses of biological variables. This algorithm is used to estimate the parameters of the two (or more) regression functions and the separation point(s) (thresholds, sensitivities) by statistical approximation. Although it is often unknown whether the response of a biological variable is adequately described by one rectilinear regression function or by piecewise linear regression function(s) with separation point(s), an *F* test is proposed to determine whether one regression line is the optimal fitted function. A FORTRAN-77 program has been developed for estimating the optimal parameters and the coordinates of the separation point(s). A few sets of data illustrating this kind of problem in the analysis of thermoregulation, osmoregulation, and the neuronal responses are discussed.

method of least squares; maximum likelihood methods; join point; change point; *F* test; confidence interval; computer program; thermoregulation; osmoregulation; activity of neurons

IN MANY BIOLOGICAL SYSTEMS the response of a variable changes at a special critical value. For instance, a thermoregulatory parameter may change little with temperature unless a critical value is transgressed above or below which the parameter will change strongly with body temperature in an approximately linear fashion. In this and similar situations the response often is described as consisting of two sections, each of which may be described by linear regression functions with different slopes. It is clear that the response of a biological variable can be described to any desirable degree of accuracy by adapting a nonlinear function (exponential function, polynomial function, or Fourier series). The parameters of these complicated functions, however, often cannot be interpreted biologically. Therefore simpler methods, such as fitting piecewise linear regression, appear more appropriate. The problem is that of determining the point of separation between the two regression functions, which is not known a priori. Usually this is done by inspection. As a more accurate approximation, the set of data may be separated at various  $x$  values, and after calculating regression functions for the various subsets, the best fit may be chosen as the point of separation. On principle, however, it should be possible to estimate this “join point” of the two regression lines by statistical methods.

This problem was first studied by Quandt (9), who

proposed a maximum likelihood method for estimating the parameters of two separate regression lines, which switch at an unknown “change point”  $x_0$ . This model did not require the assumption that the two regression lines are continuous in  $x_0$ ; i.e., the two regression lines do not meet at this point. Sprent's model (12) included as a strong assumption that  $x_0$  was known and discussed several sets of two-phase regression data with application in biometry. Hudson (6) considered the problem of the continuity constraint being included in his model (the two regression functions meet at the point  $x_0$ , the join point). He obtained the estimated parameters using the method of least squares. Hasselblad et al. (3) applied Hudson's method to the specific case of a hockey-stick function, which assumes that the first section of the data set is fitted by a linear regression with slope zero.

Although for many sets of biological data continuity may be presumed, it is often unknown whether they are adequately described by one rectilinear regression function or by two (or more) and whether, in the latter case, the assumption of a slope of zero for the first regression line is justified. Quandt (10) proposed a likelihood ratio test for two separate regression lines as opposed to the null hypothesis that the data follow only one. However, except for some special cases, the exact null distribution remains intractable, since no limiting distribution is known (4). Worsley (14) developed conservative bounds for the null distribution function of the likelihood ratio test statistic, permitting conservative tests for the assumption of two regression lines.

This paper describes statistical methods that were developed to deal with the outlined problems. Proceeding from Hudson's function, our method assesses the fit of two (or more) regression functions to a set of data without requiring a slope of zero for the first regression line. A FORTRAN-77 program has been written for estimating the parameters and the separation point of two regression functions with (join point) and without (change point) the continuity constraint. Furthermore an *F* test is proposed to test whether one regression line is the optimal fitted function.

## Methods and Results

Suppose we have  $n$  pairs of observations  $(x_i, y_i)$ ,  $i = 1, \dots, n$  and without loss of generality we assume that the  $x_i$  are ordered  $x_1 \leq x_2 \leq \dots \leq x_n$ .

The relationship between the  $x$  values and the  $y$  values

can be described by two linear regression functions of the form

$$y_1 = a_1 + b_1 \cdot x_i \quad x_i \leq x_0$$

$$y_2 = a_2 + b_2 \cdot x_i \quad x_i > x_0$$

where  $(x_0, y_0)$  represents coordinates of the critical point (change point, join point),  $a_1$  is intercept of the first regression line,  $b_1$  is slope of the first regression line,  $a_2$  is intercept of the second regression line, and  $b_2$  is slope of the second regression line.

The estimation of the parameters is based on the method of least squares. The regression line fitted by least squares is one that makes the sum of squares of all vertical deviations  $(y_i - \hat{y}_i)$  as small as possible, where  $\hat{y}_i$ ,  $i = 1, \dots, n$  are the points on the regression line belonging to the data points  $y_i$ ,  $i = 1, \dots, n$

$$\text{RSS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (a + b \cdot x_i)]^2 \rightarrow \min$$

where  $\hat{y}_i = a + b \cdot x_i$ ,  $a$  is intercept of the regression line,  $b$  is slope of the regression line, and RSS is residual sum of squares.

The parameters of the regression line are determined by differentiating the equation with respect to each parameter and setting the results to zero. This procedure leads to the so-called normal equations; the solution of these equations represents the parameters of the fitted linear regression function.

The difference between fitting one linear regression function and the method proposed in this article is as follows. As mentioned in the introduction, the response of special variables can be described by two linear regression functions with a change point (the point where there is a step from one regression line to the next) or a join point (the point where the first regression line intersects the second). These points are unknown a priori.

Without involving a primary estimation by eye, an iterative approach is used to arrive at the critical value in a finite number of steps, exclusively with exact statistical methods. With the algorithm described below, several sets of parameters are successively estimated and the optimal fit is determined using a statistical criterion.

At the first iteration the observations  $x_1, x_2$ , and  $x_3$  are included, estimating the parameters of the first regression line, and the remaining  $n - 3$  observations  $x_4, \dots, x_n$  are used fitting the second regression line. Then, at the next iteration the first regression line is fitted to the observations  $x_1, \dots, x_4$  (first part of the data), the estimation of the parameters of the second regression line is based on the remaining  $n - 4$  observations (second part of the data). In general terms, at each iteration of this algorithm, one additional observation of the second part of the data is shifted to the first part of the data; the second regression line is fitted to the remaining observations.

Then RSS is calculated, which is the sum of the residual sum of squares of each regression line

$$\text{RSS} = \sum_{x_i \leq x_0} [y_i - (a_1 + b_1 \cdot x_i)]^2 + \sum_{x_i > x_0} [y_i - (a_2 + b_2 \cdot x_i)]^2$$

RSS is used as a criterion to determine the optimal fit: the optimal values of  $a_1, b_1, a_2, b_2$ , and  $x_0$  are the values belonging to the minimal RSS.

In the introduction, we gave a short historical overview of methods pertinent for our kind of problems. In the following section we present the models and explain the methods.

The model developed first (9, 12) describes the response of a variable as consisting of two regression lines that switch at an unknown critical value, the change point. Therefore an optimal fit can be described as follows.

The change point is determined to lie between the observations  $x_j$  and  $x_{j+1}$ , and there is a step between  $x_j$  and  $x_{j+1}$ . The parameters of the first regression line are estimated from the observations  $x_1, \dots, x_j$ ; the parameters of the second regression line are based on the observations  $x_{j+1}, \dots, x_n$  (Fig. 1).

The second model (6) includes the continuity constraint; i.e., the model requires that the two regression functions meet at the join point  $x_0$  (Fig. 2).

We have to distinguish between two cases: the join point lying between the  $x$  values of two observations and the join point being identical with the  $x$  value of an observation.

Starting with the assumption that the join point is located between  $x_j$  and  $x_{j+1}$  ( $j = 3, \dots, n - 3$ ), the different steps of the algorithm are passed through including the constraint. Afterwards the same is done under the assumption that the join point is identical with  $x_j$  ( $j = 3, \dots, n - 3$ ). The algorithm differs slightly because the join point is known at each step.

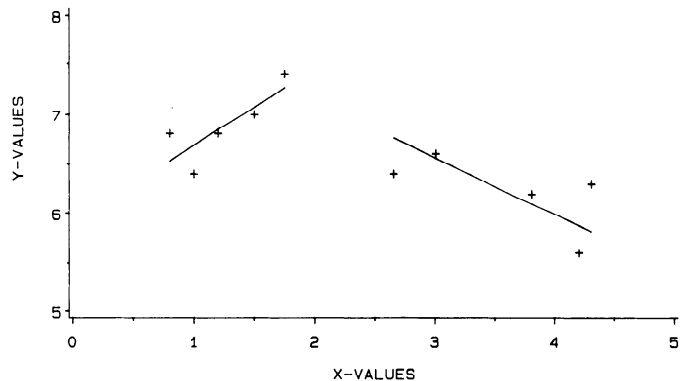


FIG. 1. Piecewise linear regression without continuity constraint.

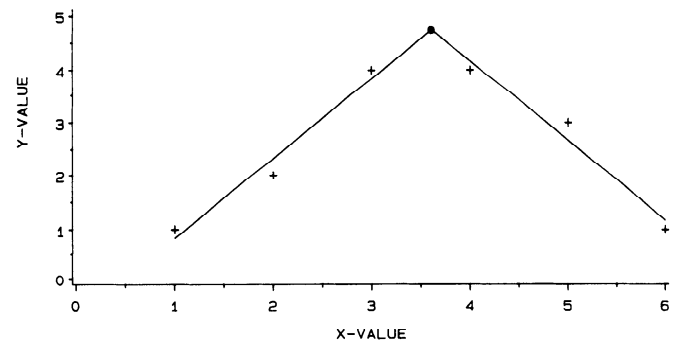


FIG. 2. Piecewise linear regression with continuity constraint.

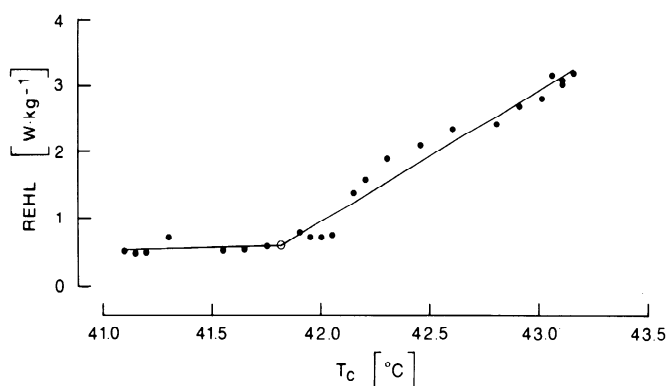


FIG. 3. Respiratory evaporative heat loss (REHL) in ducks at external heat load [1st day, ambient temperature ( $T_a$ ) = 35–38°C, relative humidity (rh) = 50%].  $T_c$ , core temperature.

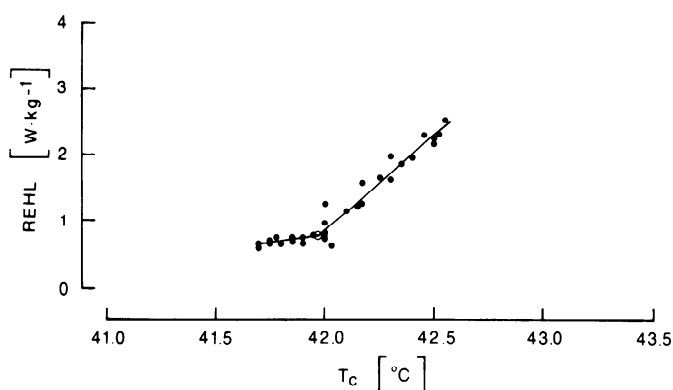


FIG. 4. REHL in ducks at external heat load (2nd day,  $T_a$  = 35–38°C, rh = 50%). See legend of Fig. 3 for definition of abbreviations.

Hasselblad et al. (3) presented exact formulas for fitting two regression functions with a join point in the case of a hockey-stick function. We expanded the exact formulas for Hudson's general model and transferred these methods to the case of three regressions with either a join point or a change point.

### Test and Confidence Interval

Under the assumption that the observations  $y_i$ , ( $i = 1, \dots, n$ ) are normally distributed, one can use methods like the  $F$  test of the linear hypothesis (11) to decide whether to include certain terms in the model.

hypothesis: is a linear regression line the optimal fitted function?

$$H_0: a_1 = a_2 \quad \text{and} \quad b_1 = b_2$$

$$\text{test statistic: } F = \frac{(\text{RSSL} - \text{RSS})/3}{\text{RSS}/(n - 4)}$$

where RSSL is residual sum of squares of linear regression; RSS is residual sum of squares of piecewise linear regression; degrees of freedom:  $n - 4$ ;  $(\text{RSSL} - \text{RSS})$ : degrees of freedom: 3.

The  $F$  test statistic belongs to the group of likelihood ratio test statistics; therefore we can apply the results of Hinkley (5), who states that when the continuity con-

straint is included in the model, the null distribution of the likelihood ratio test is asymptotically an  $F$  distribution with 3 and  $n - 4$  degrees of freedom.

If one does not want to assume continuity, the approach of Beckman and Cook (1) is recommended. They propose using critical values obtained by simulation, and for extreme configurations of the data the Bonferroni bounds are recommended.

If  $H_0$  is rejected, conventional criteria can be applied to the single regression lines with regard to their statistical significance, e.g., whether the slope of one of the regression lines is equal to zero.

Let

$$y_1 = a_1 + b_1 \cdot x_i \quad x_i \leq x_0$$

$$y_2 = a_2 + b_2 \cdot x_i \quad x_i > x_0$$

be the two fitted linear regressions based on the  $j$  observations of the first part of the data and the  $(n - j)$  observations of the second part of the data and  $x_0$  the join point where the two lines meet.  $y_k$ ,  $k = 1, 2$  are the observations of the random variables  $Y_k$ ,  $k = 1, 2$ , which are normally and independently distributed with means

$$E(Y_k) = a_k + b_k \cdot x$$

and common variance  $\sigma^2$ .

On the basis of these assumptions, Kastenbaum (7) proposed a formula for constructing a  $100(1 - \alpha)\%$

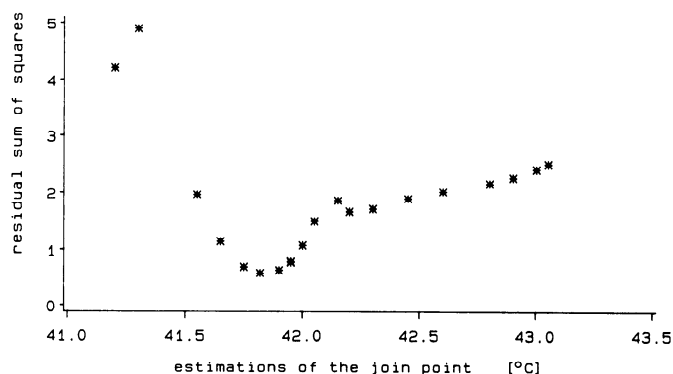


FIG. 5. Residual sum of squares as a function of estimated join points. Determination of threshold in thermoregulation (1st day). For plot of data and fitted responses, see Fig. 3.

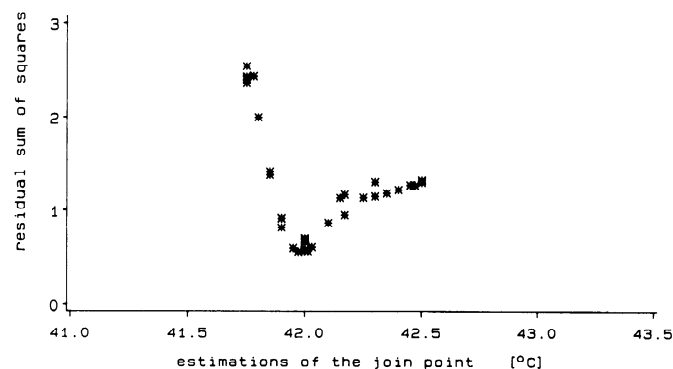


FIG. 6. Residual sum of squares as a function of estimated join points: determination of threshold in thermoregulation (2nd day). See Fig. 4 for data and regression lines with estimated join point.

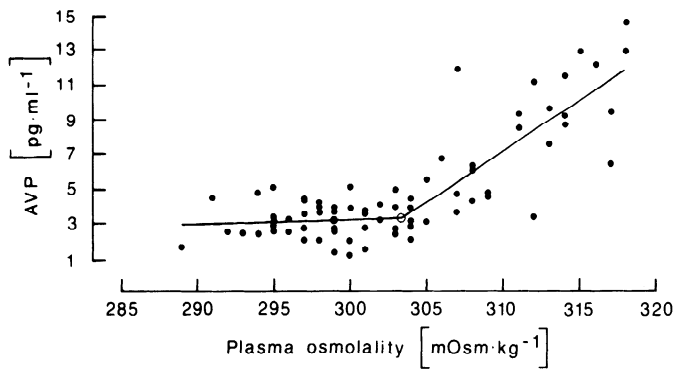


FIG. 7. Relationship between plasma osmolality and arginine vasopressin (AVP) concn in plasma of conscious dogs (determination of osmotic threshold).

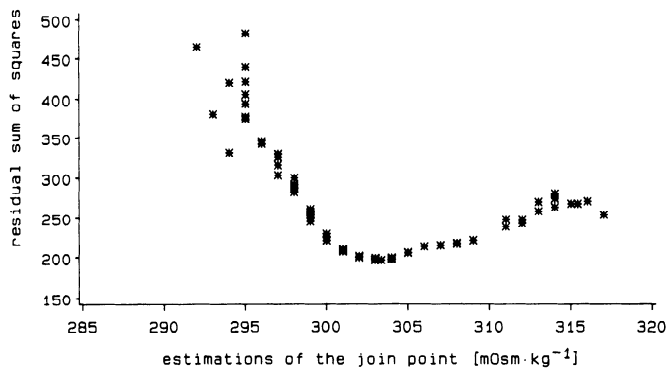


FIG. 8. Residual sum of squares as a function of estimated join points: determination of osmotic threshold. See Fig. 7 for data and fitted responses.

confidence interval for the join point,  $x_0$ , which is given by the roots of

$$[a_1 - a_2 + x_0 \cdot (b_1 - b_2)]^2 - F_{1,n-4} \cdot S^2 \cdot w = 0$$

where  $S^2$  is  $\text{RSS}/(n-4)$  estimation of  $\sigma^2$ ,  $F_{1,n-4}$  is critical value of the  $F$  distribution with 1 and  $n-4$  degrees of freedom, and

$$w = (1/j) + \sum_{i=1}^j (\bar{x}_1 - x_0)^2 / \sum_{i=1}^j (x_{1i} - \bar{x}_1)^2 + [1/(n-j)] + \sum_{i=j+1}^m (\bar{x}_2 - x_0)^2 / \sum_{i=j+1}^n (x_{2i} - \bar{x}_2)^2$$

Seber (11) published the same formulas for finding a confidence interval in his book and transformed the equation in such a way that the formula for solving quadratic equations can easily be used.

Hinkley (5) summarized the maximum likelihood estimation procedures for estimating the parameters and provided approximative confidence intervals for the join point  $x_0$  using the appropriate likelihood ratio statistics. The properties of the procedures discussed in his article are based on large-sample theory and might not be accurate in small samples; the distribution of  $x_0$  converges very slowly to normality as illustrated in the article.

## Discussion

The algorithm presented above enables us to estimate the optimal parameters and the join point(s) within a set of data by statistical approximation. The test statistic permits determination of the error probability for rejection of the null hypothesis that the set of data is described appropriately by only one regression function. In contrast to the hockey-stick function (6), the proposed method avoids setting the slope of one of the regression lines equal to zero, which does not seem appropriate a priori in describing many of the relationships to be expected between biological variables. This is true for at least some of the problems that have arisen in the analysis of thermoregulation, osmoregulation, and the neuronal responses for which the described method was developed. A few sets of data pertinent for this kind of problem and the applicability of the described method are discussed below.

**Thermoregulation.** Figures 3 and 4 show respiratory evaporative heat loss (REHL) due to thermal tachypnea (panting) as functions of core temperature in a duck. Experiments *a* and *b* were done on different days (data from Ref. 2).

The plot of the data in Fig. 3 shows that the variable REHL is nearly independent of temperature up to a

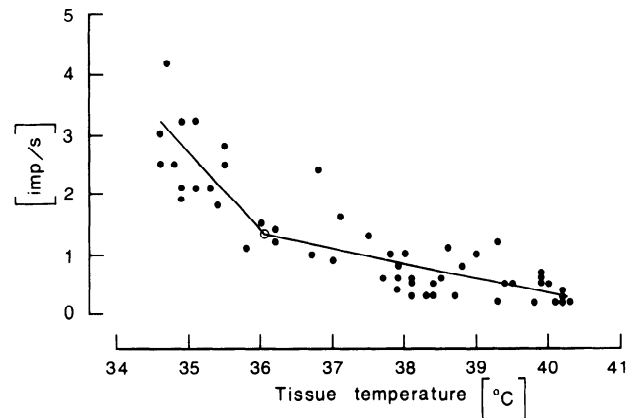


FIG. 9. Estimation of join points in nonrectilinear thermal responses: hypothalamic unit activity in vitro as a function of temperature.

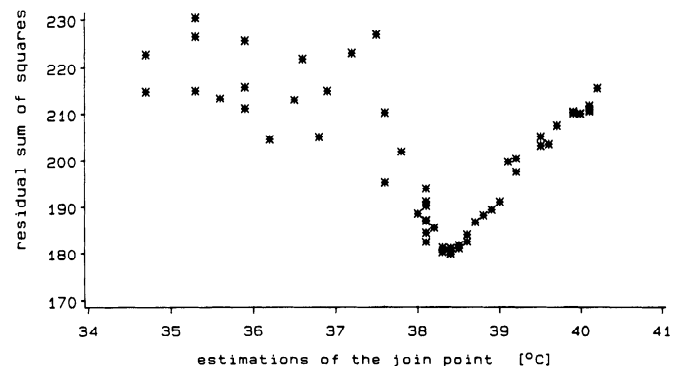


FIG. 10. Residual sum of squares as a function of estimated join points: estimation of join points in nonrectilinear thermal responses. See Fig. 9 for plot of data and fitted responses.

critical value; for the data to the right of this point, a linear correlation can be seen. We estimated the parameters of the regression lines and the join point first under the assumption of a hockey-stick function and second without limiting assumption (Fig. 5).

Including the assumption that the slope of the first regression line is zero, the coordinates of the estimated join point are  $(x_o, y_o) = (41.8, 0.57)$ ; the estimation of the slopes of the regression lines is  $(b_1 = 0)$ ,  $b_2 = 1.98$ ; the value of the calculated RSS is 0.592.

Using the model without assumption (general model) the coordinates of the join point and the estimates of the parameters differ slightly:  $b_1 = 0.072$ ,  $b_2 = 1.98$  ( $a_1 = -2.42$ ,  $a_2 = -82.13$ ); the coordinates of the join point are  $(x_o, y_o) = (41.82, 0.60)$ . The value of the RSS is a little smaller than the previous one: RSS = 0.590.

In Fig. 4, application of the same procedure to a set of data obtained from the same animal on a different day shows, as a biological result, that the threshold of the panting response  $[(x_o, y_o) = (41.97, 0.78)]$  is a comparatively stable biological parameter. The estimation of the parameters of the regression lines is as follows:  $a_1 = -17.05$ ,  $b_2 = 0.43$ ,  $a_2 = -120.83$ ,  $b_2 = 2.9$  (Fig. 6).

TABLE 1. List of data

X	Y	X	Y
289.00	1.73	301.00	3.50
291.00	4.50	301.00	2.77
292.00	2.54	302.00	3.17
293.00	2.44	302.00	4.04
294.00	2.44	303.00	4.96
294.00	4.80	303.00	3.92
295.00	3.28	303.00	2.71
295.00	5.11	303.00	2.34
295.00	2.90	304.00	2.80
295.00	3.17	304.00	4.43
295.00	2.53	304.00	3.91
295.00	2.67	304.00	3.10
295.00	3.21	304.00	2.10
296.00	2.50	305.00	5.53
296.00	3.14	305.00	3.10
297.00	3.55	306.00	6.70
297.00	2.07	307.00	11.80
297.00	4.51	307.00	3.63
297.00	2.75	307.00	4.71
297.00	4.30	308.00	6.00
298.00	2.04	308.00	4.30
298.00	4.20	308.00	6.23
298.00	2.01	309.00	4.53
298.00	3.86	309.00	4.79
298.00	3.58	311.00	8.40
299.00	2.67	311.00	9.21
299.00	2.55	312.00	11.12
299.00	3.70	312.00	3.40
299.00	2.71	313.00	7.50
299.00	3.60	313.00	9.60
299.00	1.41	314.00	9.20
299.00	3.96	314.00	8.60
299.00	3.21	314.00	11.50
300.00	1.26	315.00	12.80
300.00	3.92	316.00	12.10
300.00	2.06	317.00	6.40
300.00	5.10	317.00	9.40
301.00	1.53	318.00	14.50
301.00	3.67	318.00	12.80

X, plasma osmolality; Y, arginine vasopressin concn in plasma.

TABLE 2. Output of FORTRAN program

Piecewise Linear Regression
Minimum of RSS
(A) Under the assumption: $X_0$ in $[X(J), X(J+1)]$
Min(SQERR1) = 196.381
(B) Under the assumption: $X_0 = X(J)$
Min(SQERR2) = 196.948
Minimum of RSS of A and B
Min(SQERR1, SQERR2) = 196.381
Parameter of optimal fit
(1) Regression line:
Intercept A1 = -2.501
Slope B1 = 0.019
(2) Regression line:
Intercept A2 = -174.711
Slope B2 = 0.587
Joint point of regression lines:
$(X_0, Y_0) = (303.371, 3.267)$
Linear Regression
RSS = 286.427
Parameter
Intercept (A) = -96.303
Slope (B) = 0.334
Hypothesis: "Linear Regression Is the Optimal Fit?"
F test
$F = 16.9654$ $F(3,74)$ —distributed
$P = 0.46963292E-07$

**Osmoregulation.** Figure 7 shows arginine vasopressin (AVP) concentration in plasma as a function of plasma osmolality in conscious dogs (data from Ref. 13).

In a previous evaluation, the  $x$  values between 295 and 305 had been tentatively tested as the threshold for the osmotic stimulation of plasma AVP concentration. The best fit by linear regression was found with 301 mosmol·kg<sup>-1</sup>. With the algorithm proposed in this study, the best fit with the largest values of the coefficients of correlation and the smallest RSS were obtained with the join point  $(x_o, y_o) = (303.37, 3.27)$ . (Figure 8 shows the RSS as a function of the estimated join points.)

**Activity of neurons.** Figure 9 presents discharge frequencies of single neurons as a functional local temperature recorded extracellularly in vitro from hypothalamic slices (data from 8) (Figs. 9 and 10).

The plot of the data shows a generally positive relationship between the two variables. The activity of neurons changes little with temperature unless a critical value is transgressed above which the activity changes strongly with temperature. The estimated slopes of the regression lines are  $b_1 = 0.626$ ,  $b_2 = 2.447$ . The coordinates of the join point are  $(38.4, 3.44)$ , where the  $x$  coordinate of the join point is identical with the  $x$  value  $x_{37}$  of the observations. Figure 10 presents the RSS as a function of the estimated join points.

Because there is no point fitting a line to less than 3 data points the iterations start at  $j = 3$  (first iteration: the  $x$  values  $x_1, x_2, x_3$  are included fitting the first regression line) and stop when  $j = n - 3$  is reached (last iteration: the  $x$  values  $x_{n-2}, x_{n-1}, x_n$  are used to estimate the parameters of the second regression line). If the first  $m$  ( $m > 2$ ) observations have the same  $x$  value, the iterations start later. The first iteration includes the  $x$  values  $x_1, \dots, x_{m+1}$  fitting the first regression line. The  $x$

values  $x_{m+2}, \dots, x_n$  are included estimating the second regression function. where

If the last  $m$  ( $m > 2$ ) observations have the same  $x$  value, the  $x$  values  $x_{n-m}, \dots, x_n$  are included estimating the parameters during the last iteration [the first regression line is fitted to  $x_1, \dots, x_{n-(m+1)}$ ].

### Numerical Example

The data of an experiment in osmoregulation are shown in Table 1; Fig. 7 gives an impression of the relationship of these two variables [plasma osmolality ( $X$ ) and AVP concentration ( $Y$ ) in conscious dogs].

The minimal RSS and the parameter estimations calculated by the computer program (for the piecewise linear regression and for the linear regression) are displayed in Table 2.

The value of the test statistic exceeds the critical value of the  $F$  distribution with 3 and 74 degrees of freedom. Therefore the hypothesis that the data can be described by one rectilinear function must be rejected.

### Program

Join or change point and the parameters of the two or three regression functions are estimated with a FORTRAN-77 program. The graphical presentation of data, critical point and fitted regression lines, is done with SAS/GRAPH (Statistical Analysis System) on a plotter (Hewlett-Packard model HP-7550A). Both programs run on the VAX 11/780 (Digital Equipment).<sup>1</sup>

## APPENDIX

### Summary of the Algorithm

#### 1. Piecewise linear regression

function

$$y_{1i} = a_1 + b_1 \cdot x_i \quad x_0 \leq x_i$$

$$y_{2i} = a_2 + b_2 \cdot x_i = y_0 + b_2 \cdot (x_i - x_0) \quad x_0 > x_i$$

where  $a_1, a_2$  is intercept,  $b_1, b_2$  is slope, and  $(x_0, y_0)$  are coordinates of the join point.

restriction

$$a_1 + b_1 \cdot x_0 = a_2 + b_2 \cdot x_0$$

#### estimation of the parameters

a) assumption:  $x_0$  is situated between  $x(j)$  and  $x(j+1)$  ( $j = 3, \dots, n-3$ )

$$b_1(j) = \frac{\sum_{i=1}^j x_i \cdot y_i - \frac{1}{j} \cdot \sum_{i=1}^j x_i \cdot \sum_{i=1}^j y_i}{\sum_{i=1}^j x_i^2 - \frac{1}{j} \cdot \left( \sum_{i=1}^j x_i \right)^2}$$

$$a_1(j) = \bar{y}_1(j) - b_1(j) \cdot \bar{x}_1(j)$$

$$x_0(j) = \frac{\sum_{i=j+1}^n y_i \cdot \sum_{i=j+1}^n x_i^2 - \sum_{i=j+1}^n x_i \cdot y_i \cdot \sum_{i=j+1}^n x_i + a_1(j) \cdot hf(j)}{\sum_{i=j+1}^n x_i \cdot \sum_{i=j+1}^n y_i - (n-j) \cdot \sum_{i=j+1}^n x_i \cdot y_i - b_1(j) \cdot hf(j)}$$

$$hf(j) = \left( \sum_{i=j+1}^n x_i \right)^2 - (n-j) \cdot \sum_{i=j+1}^n x_i^2$$

$$y_0(j) = a_1(j) + b_1(j) \cdot x_0(j)$$

$$b_2(j) = \frac{\sum_{i=j+1}^n y_i - (n-j) \cdot y_0(j)}{\sum_{i=j+1}^n x_i - (n-j) \cdot x_0(j)}$$

$$a_2(j) = y_0(j) - b_2(j) \cdot x_0(j)$$

b) assumption:  $x_0 = x(j)$ , the  $x$  coordinate of an observation [ $j = 3, \dots, (n-3)$ ]

$a_1(j), b_1(j)$ : estimation analogous to (a)

$$x_0(j) = x(j)$$

$$y_0(j) = a_1(j) + b_1(j) \cdot x_0(j)$$

$$b_2(j) = \frac{\sum_{i=j+1}^n y_i - (n-j) \cdot y_0(j)}{\sum_{i=j+1}^n x_i - (n-j) \cdot x_0(j)}$$

$$a_2(j) = \frac{\sum_{i=j+1}^n x_i \cdot y_i}{\sum_{i=j+1}^n x_i} - b_2(j) \cdot \frac{\sum_{i=j+1}^n x_i^2}{\sum_{i=j+1}^n x_i}$$

residual sum of squares

$$\begin{aligned} \text{RSS}(j) = & \sum_{x_i \leq x_0} \{y_i - [a_1(j) + b_1(j) \cdot x_i]\}^2 \\ & + \sum_{x_i > x_0} \{y_i - [a_2(j) + b_2(j) \cdot x_i]\}^2 \end{aligned}$$

$$\text{RSS} = \min_j \text{RSS}(j) \quad j = 3, \dots, n-3$$

with the parameter estimates  $a_1, b_1, a_2$ , and  $b_2$ .

Under the assumption  $x_0$  is situated between  $x(j)$  and  $x(j+1)$ , the estimate  $x_0(j)$  can sometimes fall outside the interval that divides the data set. In this case, we put  $\text{RSS}(j) = \infty$  to exclude this RSS from choosing the minimum RSS (6).

#### 2. Simple linear regression

estimation of parameters

$$\begin{aligned} a &= \bar{y} - b \cdot \bar{x} \\ b &= \frac{\sum_{i=1}^n x_i \cdot y_i - \frac{1}{n} \cdot \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \cdot \left( \sum_{i=1}^n x_i \right)^2} \end{aligned}$$

residual sum of squares

$$\text{RSSL} = \sum_{i=1}^n [y_i - (a + b \cdot x_i)]^2$$

#### 3. F test

test statistic

$$F = \frac{(\text{RSSL} - \text{RSS})/3}{\text{RSSL}/(n-4)}$$

distribution of the test statistic

$F$  distribution with 3 and  $n-4$  degrees of freedom.

<sup>1</sup> A source listing of the programs is available from the author.

Received 1 June 1988; accepted in final form 7 March 1989.

## REFERENCES

1. BECKMAN, R. J., AND R. D. COOK. Testing for two-phase regression. *Technometrics* 21: 65-69, 1979.
2. GEISTHÖVEL, E., O. LUDWIG, AND E. SIMON. Capsaicin fails to procedure disturbances of autonomic heat and cold defence in an avian species (*Anas platyrhynchos*). *Pfluegers Arch.* 406: 343-350, 1986.
3. HASSELBLAD, V., J. P. CREASON, AND W. C. NELSON. *Regression Using "Hockey-Stick" Functions*. Research Triangle Park, NC: US Environmental Protection Agency, 1976.
4. HAWKINS, D. M. A note on continuous and discontinuous segmented regression. *Technometrics* 22: 443-444, 1980.
5. HINKLEY, D. V. Inference in two-phase regression. *J. Am. Stat. Assoc.* 66: 736-743, 1971.
6. HUDSON, D. J. Fitting segmented curves whose points have to be estimated. *J. Am. Stat. Assoc.* 61: 1097-1129, 1966.
7. KASTENBAUM, M. A. A confidence interval on the abscissa of the point of intersection of two fitted linear regressions. *Biometrics* 15: 323-324, 1959.
8. NAKASHIMA, T., F.-K. PIERAU, E. SIMON, AND T. HORI. Comparison between hypothalamic thermoresponsive neurons from duck and rat slices. *Pfluegers Arch.* 409: 236-243, 1987.
9. QUANDT, R. E. Maximum-likelihood method for estimating the parameters in two separate regression lines that switch at an unknown point  $k$ . *J. Am. Stat. Assoc.* 53: 873-880, 1958.
10. QUANDT, R. E. A likelihood ratio test to test two separate regression lines as opposed to the null hypothesis that the data follow only one. *J. Am. Stat. Assoc.* 55: 324-330, 1960.
11. SEBER, G. A. F. *Linear Regression Analysis*. New York: Wiley, 1977, p. 96-108, 197-209.
12. SPRENT, P. Some hypothesis concerning two-phase regression lines. *Biometrics* 17: 634-645, 1961.
13. SZCZEPANSKA-SADOWSKA, E., D. GRAY, AND C. SIMON-OPPERMANN. Vasopressin in blood and third ventricle CSF during dehydration, thirst and hemorrhage. *Am. J. Physiol.* 245 (*Regulatory Integrative Comp. Physiol.* 14): R549-R555, 1983.
14. WORSLEY, K. J. Testing for a two-phase multiple regression. *Technometrics* 25: 35-42, 1983.

